Training Workers for Improving Performance in Crowdsourcing Microtasks

Ujwal Gadiraju[1], Besnik Fetahu[1] and Ricardo Kawase[2]

[1] L3S Research Center, Leibniz Universität Hannover,
30167 Hannover, Germany
{gadiraju, fetahu}@L3S.de
[2] mobile.de, Marktplatz 1, 14532 Europarc Dreilinden, Germany
rkawase@team.mobile.de

Abstract. With the advent and growing use of crowdsourcing labor markets for a variety of applications, optimizing the quality of results produced is of prime importance. The quality of the results produced is typically a function of the performance of crowd workers. In this paper, we investigate the notion of treating crowd workers as '*learners*' in a novel learning environment. This learning context is characterized by a short-lived learning phase and immediate application of learned concepts. We draw motivation from the desire of crowd workers to perform well in order to maintain a good reputation, while attaining monetary rewards successfully. Thus, we delve into training workers in specific microtasks of different types. We exploit (i) *implicit training*, where workers are provided training when they provide erraneous responses to questions with priorly known answers, and (ii) *explicit training*, where workers are required to go through a training phase before they attempt to work on the task itself. We evaluated our approach in 4 different types of microtasks with a total of 1200 workers, who were subjected to either one of the proposed training strategies or baseline case of *no training*. The results show that workers who undergo training depict an improvement in performance upto 5%, and a reduction in the task completion time upto 41%. Additionally, crowd training led to the elimination of malicious workers and a costs-benefit gain upto nearly 15%.

Keywords: Crowdsourcing · Workers · Training · Learning · Microtask · Performance

1 Introduction

In the recent years, crowdsourcing has become a convenient instrument for solving a multitude of problems that require human input. The methodology of crowdsourcing encompasses the 'outsourcing' of work to the crowd, where each crowd worker performs a small part of a task, commonly known as a 'microtask'. By accumulating the responses from workers through such 'microtasks', requesters solve different problems.

Crowdsourcing has been employed in many areas; to assemble dictionaries [3], to map the outer space [4], to aid in disaster relief [20], or in a most recent case, to support

[3] http://www.wired.co.uk/news/archive/2011-01/13/the-oxford-english-wiktionary
[4] http://edition.cnn.com/2014/08/17/tech/nasa-earth-images-help-needed/

the search of a missing flight [5]. Wikipedia is a remarkable example of crowdsourcing that is well known to internet users. Millions of people contribute with small pieces of information to build a complete knowledge base. Lately, researchers in several fields are extensively employing crowdsourcing to collect human annotated data, build ground truths, to evaluate and validate results.

In this paper, our main goal is to tackle the aforementioned challenges that result in sub-optimal performance from crowd workers, and evaluate the best means to improve workers performance. Here we recognize the crowd worker as a learner, who is placed in an atypical learning set up that differs from traditional learning environments. First, there is no background information on the knowledge, skills or competences of the workers. Second, given the nature of the microtasks (short and less time-consuming), workers face an 'on-the-fly' learning situation. It is comparable to '*Experiential learning*' (the process of learning through experience, or learning by doing), where workers acquire experience during the execution of the task. It is also comparable to '*Microlearning*', which deals with small learning units and short-term learning activities. However, in many cases of microtasks, workers have no time to apply their gained experience. In some cases, workers who are performing badly throughout a task are not allowed to continue, while in other cases, workers are not aware of their performance. Thirdly, workers are highly motivated to perform tasks accurately, in order to receive the monetary reward. Finally, in this set up workers learn through training in a non-collaborative environment.

Taking into consideration this very particular learning setup, in this paper, we evaluate two different learning strategies that support the crowd workers to achieve high-quality performance, namely *implicit training* and *explicit training*.

Definition 1. Implicit Training. Workers are prompted with training information according to the microtask at hand, when they provide erroneous responses to questions with priorly known answers.

Definition 2. Explicit Training. Workers are required to complete a training session where they are prompted with training information according to the microtask at hand and asked to complete sample microtasks, before they begin responding to the questions within the task.

We deployed 12 crowdsourced tasks, spanning a total of 800 crowd workers who were subjected to one of these training strategies, and 400 workers who were not provided with any training at all. The main contributions of our research are listed below.

1. We provide an analysis on the learning behavior of crowd workers.
2. We identify the best training strategy for crowdsourcing microtasks that optimize the quality of results produced.
3. We compare the qualitative affects of training on different types of microtasks.
4. We show that training workers can reduce the time required for task completion, deter the participation of malicious workers and improve the overall result quality.

In addition to these tangible contributions, we believe that we are delineating a new learning process that has not been investigated thoroughly or tackled, opening space for future research in the field. The main characteristics of this learning set up is that it is short term, on-the-fly, and typically for single use.

[5] `http://tinyurl.com/kl3mmme`

2 Related Literature

In this section, we review literature related to our work. First we motivate the direction of our work in this paper. Next, we present a discussion that is mainly centered around training crowd workers in virtual environments, and compares approaches that tackle a similar problem in a crowdsourcing setup.

In the crowdsourcing paradigm, motivation plays a big role in getting the crowd engaged. A sense of participation and community, reputation and altruism are general incentives that lead people to participate and contribute in crowdsourced initiatives. However, monetary incentive is a significant factor that motivates crowd workers to engage, especially in online setups. Our recent research work shows that monetary reward is in fact, the major incentive for crowd workers to perform these microtasks [4], and finding the right balance between *effort* required to solve a task and reward is a common problem faced by task administrators[6].

Much has been done towards improving crowdsourcing methodologies, in terms of increasing workers motivation and the quality of results [7,14], improving throughput [9], and reduction of task abuse [5]. In all these cases, the common goal was to optimize the outcomes of the crowdsourcing initiatives in terms of quantity and quality.

In the pursuit of monetary rewards, workers commonly perform tasks for which they do not possess the right competence or skills. Nevertheless, these workers in many cases ignore these limitations and still attempt to perform a given microtask. In some cases, as established in our recent work, *ineligible workers* intentionally ignore restrictive instructions [5], and in other cases, the workers are not aware of their lack of competence. Poor responses can also be a result of the lack of experience of workers or familiarity with microtasks and crowdsourcing platforms. As a consequence, workers provide sub-optimal input, and final outcomes may not be fully reliable. This leads to a waste of time and resources by both parties involved.

Ensuring Quality of Crowdsourced Results. Many prior works have focused on improving worker accuracy and the quality of the responses produced by crowd workers [7,13]. A common approach to ensure the reliability of answers is by collecting multiple judgements from the crowd, such that the correct answer is determined via simple majority voting [16]. In our previous work, we have proposed tackling the challenge of improving the quality of crowdsourced results by preventing the participation of spammers and malicious workers on one hand [5], while improving microtask desgin on the other hand [12]. In contrast to such exisiting approaches, we propose training crowd workers within the specific microtask enviroment in order to improve their performance and the overall quality of the results produced. We experimentally investigate this approach on various types of microtasks.

Characterizing Crowds and Microtasks In the work done by Lin et al. [11], authors tackled the problem of improving the accuracy of answers given by the crowd workers on open-ended questions (e.g. *find the mobile number of a CEO*). By modeling the workers and the microtasks as a generative probabilistic model, they inferred correct answers, hence reducing the number of answers needed in contrast to typical majority

[6] A person who deploys a microtask in order to gather responses from the crowd. Also called a 'requester'.

voting schemes. In a similar setting, Karger et al. [8] proposed an optimization model with given constraints such as the reliability and monetary cost of a task. The main goal of this approach was to achieve an acceptable reliability of a task with a given budget constraint. Welinder et al. [17] additionally introduced the notions of difficulty of a task and the skill-set of workers. By doing so, they grouped workers that have a certain skill-set and hence produce more reliable answers corresponding to the task difficulty. Whitehill et al. [18] proposed an approach that goes one step further and automatically infers labels for a given microtask, outperforming the majority voting mechanism. The difference between the aforementioned works and our work in this paper are twofold. We consider a variety of tasks that go beyond multiple choice questions, and through training the workers we show that we can improve the reliability of answers provided by the crowd. Hence, previous approaches that can be used to improve existing baselines such as *majority voting*, are rather complementary to our approach.

Crowd Training - *Workers* as *Learners* One of the first works on teaching crowds by Goldman and Kearns [6], define the *teaching dimension*. It deliberates on the minimum number of examples needed to teach a given concept. The teaching problem is a binary classification problem, where a given hypothesis for teaching is accepted or rejected with respect to a given concept. The difference when compared to our *explicit* and *implicit* crowd training, is that we engage the crowd workers through guidelines, hints and tips that go beyond mere instructions (for example providing details on how to find the middle-name of an actor). With such training, we find that crowd workers can succesfully complete tasks at faster rates. This is all the more important, since typically crowdsourced microtasks [4] go well beyond the scope of binary answers.

In [1], the authors analyzed the task performance and the learning outcomes on a real-world classroom setup. They showed an improvement in the task performance of students subjected to appropriate learning conditions. Conversely, crowdsourcing has been used to improve the learning process in classrooms [2,19]. The main usage is primarly for obtaining feedback on the learning material and other classroom work. In our setup we train the workers to use appropriate tools and strategies to solve the tasks under evaluation adequately. Le et al. [10] investigated the effect of training questions distribution on the performance of crowd workers. They studied the relevance categorization task, and showed that a uniform distribution of training within a task produces optimal peaks in worker performance. In another closely related previous work by Singla et al. [15], the authors propose *teaching with partial feedback* for binary classification problems. The approach is iterative and at each iteration a learner is required to guess the label of an example. The learning process is influenced by showing the true label of the example. While we adopt a similar training process, we additionally investigate the effect of training on various types of microtasks.

3 Approach

In this section, we describe the methodology that we have adopted to gather data and design tasks of different types. We aim to investigate the following research questions:

– **RQ#1**: Can *training* workers on carrying out particular microtasks successfully lead to an improvement in their task performance?

– **RQ#2**: What kind of *training* results in an improvement in the quality of the results produced?
– **RQ#3**: How do *training processes* qualitatively affect different types of microtasks in crowdsourcing systems?

3.1 Methodology

We consider different microtask types according to the microtask taxonomy presented in our previous work [4]. The top-level microtask types in the proposed taxonomy are : *Information Finding*, *Verification and Validation*, *Interpretation and Analysis*, *Content Creation*, *Surveys*, and *Content Access*. Due to the fact that tasks of the type *Surveys* are subjective and questions in surveys predominantly reflect opinions, we do not consider crowd training for such microtasks. Similarly, *Content Access* tasks typically require workers to merely consume certain content, and we do not consider this type of microtask for training either.

We deploy the following tasks, to gather data representing the different types of microtasks that we aim to investigate.

– Finding the middle-names of famous people (Information Finding-**IF**)
– Spam detection (Verification and Validation-**VV**)
– Sentiment Analysis (Interpretation and Analysis-**IA**)
– Image transcription (Content Creation-**CC**)

3.2 Tasks Design and Particulars

We consider three different setups for each of the four types of tasks: implicit training (**IT**), explicit training (**ET**) and no-training (**NT**). We use a premier crowdsourcing platform, CrowdFlower[7], to deploy the 12 microtasks. We gather responses from 100 independent crowd workers in each case. On average, we offer a monetary incentive of 5 USD cents for each worker that successfully completes a task. In each task, the order of the questions are randomized for different workers, to avoid any bias possibly introduced due to specific order of the questions. In addition, clear instructions are provided to the workers for each task, in order to avoid bias due to the misinterpretation of the tasks. Lengths of the different tasks were designed such that the required completion times are similar, in order to make sound comparisons.

3.3 Training Crowd Workers

In our work, we aim to investigate the affect of *Implicit* and *Explicit Training* on the performance of crowd workers with respect to different types of microtasks.

[7] `http://www.crowdflower.com/`

Information Finding In case of the *IF* tasks, workers are asked to find the middle-names of famous people. There are 20 units in each task. For the *implicit training*, on committing errors workers are trained with the workflow of finding accurate middle-names by leveraging proper search mechanisms as shown in Figure 1. For the task with *explicit training*, workers are provided with similar training on 5 units before they begin working on the actual task.

Find the middle-name of John Kennedy.

Error Reason:
The middle-name of John Kennedy is 'Fitzgerald'.
Please use Google Search to attain help with finding the middle-names.
For e.g., by entering 'John Kennedy' on Google Search (`http://google.com`) ,
you will find links to Webpages that can give you the required information.
You can also search on Wikipedia (`http://en.wikipedia.org/`).

Fig. 1: Training in the IF task of finding middle-names (the question is inset here).

Verification and Validation In case of the *VV* tasks, workers are asked to detect the spam SMS from a set of 5 SMS texts in each of 30 units, as shown in the Figure 2. Workers are trained in the *implicit* and *explicit* training tasks with examples that emulate the following: 'fake advertisements or meaningless offers are good candidates of being spam messages'.

Please select the SPAM SMS message from the given SMS messages.
- Just forced myself to eat a slice. I'm really not hungry tho. This sucks. Mark is getting worried. He knows I'm sick when I turn down pizza. Lol
- Lol your always so convincing.
- Did you catch the bus ? Are you frying an egg ? Did you make a tea? Are you eating your mom's left over dinner ? Do you feel my Love ?
- Congrats! 1 year special cinema pass for 2 is yours. call 09061209465 now! C Suprman V, Matrix3, StarWars3, etc all 4 FREE! bx420-ip4-5we. 150pm. Dont miss out!
- I'm back & we're packing the car now, I'll let you know if there's room

Fig. 2: Example unit in the Spam Detection (VV) task.

Interpretation and Analysis In the *IA* tasks, workers are asked to assess the sentiment expressed within tweets (`Positive`, `Negative` or `Neutral`). These tasks also consist of 30 units, such as the example portrayed in Figure 3.

Had a pretty good night. Jogging on the beach, swimming in my clothes. Popcorn and Family Guy. Excellent.
- Positive
- Negative
- Neutral

Fig. 3: Example unit in the Sentiment Analysis (IA) task.

Workers are provided training in the tasks with *implicit* and *explicit* training to identify tweets with each of the 3 sentiments. For example, an excerpt from the training is: "In order to determine the underlying sentiments of a tweet, adjectives and adverbs play a key role (as expressed in this tweet through the term, '`awesome`')."

Content Creation Finally, in case of the *CC* tasks, workers are asked to transcribe CAPTCHAs[8] in 30 units within the tasks as shown in Figure 4. As a part of the training, workers are provided with tips on recognizing subtle differences between letters to perform accurately. An example excerpt is the following: 'NOTE that the second letter is '`g`'. While it may seem similar to a '`q`', the '`g`'s have a small curve at the bottom, while the '`q`'s have an underscore at the bottom'.

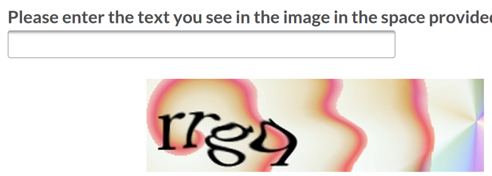


Fig. 4: Example unit in the Image Transcription (CC) task.

12 tasks were deployed as described above, with varying microtask type and training setups. Responses from workers were thereby collected in order to study the implications of training crowd workers.

4 Data Analysis

In this section, we first present details regarding the geographical distribution of crowd workers that provided responses to the different types of microtasks. Then we disuss the effect of training on various aspects.

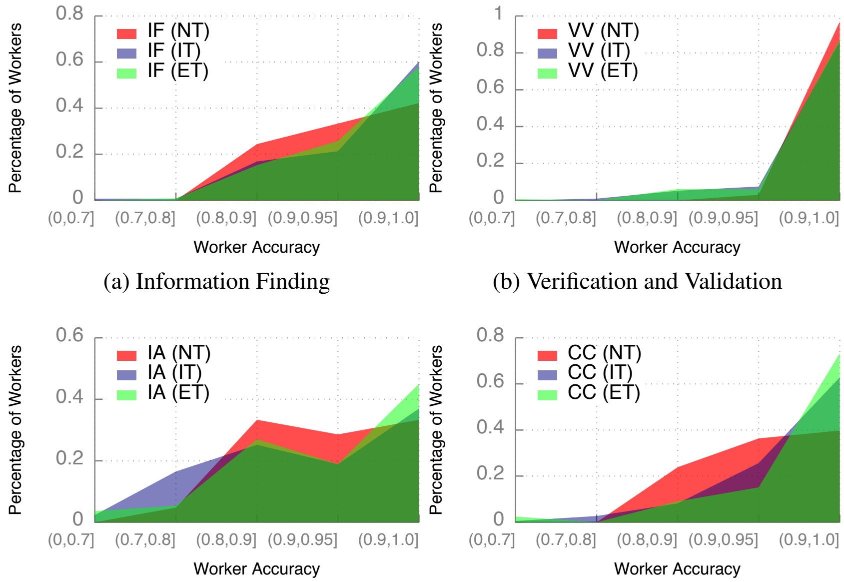
Note that, we detect *malicious workers* based on the approach described in [5], and discard the responses from these workers in our analysis, unless stated otherwise. Malicious workers are those workers who provide ill-fitting or arbitrary responses in order to complete a given task quickly and attain the monetary compensation.

We analyzed the geographical distribution of workers who participated in each of the 12 tasks deployed. Note that we did not enforce any restrictions in the participation of workers. In case of the *Information Finding* tasks we found that over 50% of the workers were from India, followed by nearly 25% workers from Turkey. Interestingly, we note that 53% of workers in case of the *Verification and Validation* tasks were from Spain, while another 24% are from Italy. For the *Interpretation and Analysis* tasks, we observed that a majority of workers (around 37%) were from Bosnia and Herzegovina, followed closely by workers from Turkey (over 31%). Finally, in case of the *Content Creation* tasks, we found that around 40% of the workers were from the Philipines and 30% of the workers were from Great Britain. We believe that these differences in the geographical distribution of the workers is primarily due to the difference in the timings at which the tasks were deployed.

4.1 Impact of Training on the Accuracy of Workers

In order to assess the impact of training on the accuracy of workers, across the different types of tasks, we analyze the responses gathered from the 12 tasks as mentioned

[8] `http://www.captcha.net/`



(a) Information Finding (b) Verification and Validation

(c) Information Analysis (d) Content Creation

Fig. 5: Accuracy of workers in different training scenarios (*No Training* (**NT**), *Implicit Training* (**IT**), and *Explicit Training* (**ET**)) for (a) IF, (b) VV, (c) IA, and (d) CC types of microtasks.

in Section 3. Figure 5 presents our findings. In Figure 5(a), we see that in the case of an Information Finding task, training helps to improve the accuracy of workers. However, there is very little difference between the increase in accuracy achieved due to *implicit* and *explicit training*. On average we find that IT and ET improve the accuracy of workers by over 2% in the IF task.

Interestingly, we find that training did not result in an improvement in the accuracy of workers in case of the Verification and Validation task ((Figure 5(b)). We note that workers achieve an accuracy of around 82% without any training. When implicit training is provided the accuracy of workers deteriorates to 65%. In case of explicit training, the accuracy of workers reduces to nearly 77%. We believe that this is due to the ease of the VV task we have considered (detection of a spam SMS), as confirmed by the average unit completion time for this particular task when compared to other tasks in the absence of training (see Figure 6b).

In the case of either an IA task (Figure 5(c)) or a CC task (Figure 5(d)), we note that training improves the accuracy of workers by around 3% and 5% on average respectively. We observe a trend where, explicit training results in the biggest improvement in the accuracy of workers, followed by implicit training. Based on our experimental findings, we observe that training workers either implicity or explicity improves their performance across all types of microtasks.

4.2 Impact of Training on Distribution of Elite Workers

As defined in our previous work [5], *elite workers* are those crowd workers that provide perfect responses to all questions within a task. We additionally analyze the affect of training workers through the distribution of *elite workers*. Figure 6a presents our findings. We observe that in the IF, IA and CC tasks, training increases the number of elite workers. In these cases, explicit training brings about the biggest increase in the percentage of elite workers, followed by implicit training. Interestingly, in case of the VV task, we observe that training decreases the number of elite workers.

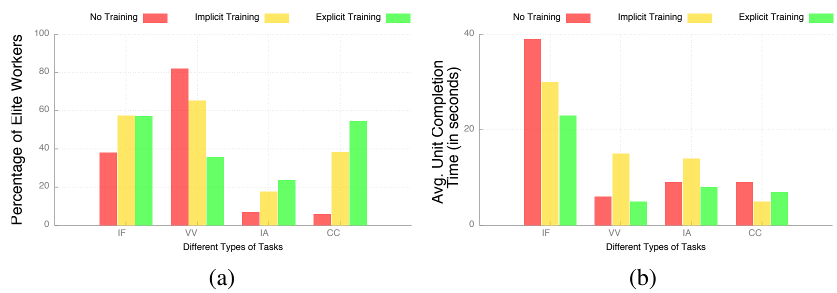


Fig. 6: Effect of training on (a) the distribution of elite workers, and (b) the average unit completion time of workers in different types of tasks.

4.3 Impact of Training on Unit Completion Time of Workers

We analyze the average time that the workers take to complete a single unit in each of the different tasks, and within different training settings. We present our findings in the Figure 6b. We find that the biggest impact of training is observed in case of the *Information Finding* task. In this task workers were trained with directions on finding middle-names in an efficient manner. We observe that such training given implicitly reduces the average *Unit Completion Time* (UCT) by 23%, and explicit training reduces the average UCT by 41%. We observe a similar trend where training decreases the average UCT of workers in case of the other task types. Interestingly, we note that implicit training increases the UCT of workers in case of the IA task.

4.4 Impact of Training on Malicious Workers and its Implication on Costs-benefit Ratio of the Responses

We additionally analyze the effect of training crowd workers with respect to the number of malicious workers attracted to a task, and present our findings in Table 1. We observe a general trend where the training process results in a decrease in the number of malicious workers. In case of the **IF** tasks, we note that implicit training (IF-IT) reduces the fraction of malicious workers from 10% to 5% and explicit training reduces this further to only 1.5%. We observe a similar trend with respect to implicit and explicit training for the tasks of type **IA** and **CC**. In case of the **VV** task, we note that while implicit training (VV-IT) seems to increase the fraction of malcious workers, explicit training reduces this by a big margin. Overall, we find that explicit training dissuades the participation of malicious workers to a large extent.

Table 1: Effect of different types of training on the fraction of malicious workers in varying types of microtasks.

Microtask Type	No Training(**NT**)	Implicit Training(**IT**)	Explicit Training(**ET**)
IF	10%	5%	1.5%
VV	2%	3.2%	0.6%
IA	37%	25.7%	22.4%
CC	12.9%	9.9%	4.2%

The effect of training, resulting in the decrease of malicious workers has important implications on the monetary compensation that task requesters have to provide. Note that the financial rewards that malicious workers attain on task completion contribute to a wastage of resources, since their responses cannot be used by a requester. Therefore, with the reduction in the fraction of malicious workers in a given task, there is a corresponding gain in the costs-benefit ratio with respect to the responses gathered. In case of **IF** tasks, we note that ET improves the costs-benefit ratio by 8.5% and IT improves the costs-benefit ratio by 5%. Similarly, we note an increase in the costs-benefit ratio by 1.4% in case of **VV** tasks, 14.6% in case of **IA** tasks, and 8.5% in case of **CC** tasks through explicit training of crowd workers.

4.5 Overall Accuracy of Tasks in Different Training Settings

We finally investigate the overall accuracy that is achieved in tasks of different types under the varying training setups. The box-plots in Figure 7 depict our findings. We observe that the overall task accuracy improves when the crowd workers are trained either implicitly or explicitly. In case of the VV tasks, we note that there is no significant improvement due to the training. As discussed earlier, we believe that this is due to the relative ease of the particular *Spam Detection* task that was considered. We clearly observe that the detection and removal of malicious workers improves the overall task accuracy across different task types, when compared to their contrasting counterparts.

5 Discussion

Through our analysis of the different types of microtasks we investigated in varying crowd training settings, we can conclude that training workers improves the accuracy of the workers, thereby enhancing the overall quality of the results produced. An important implication of *implicit* and *explicit training* of crowd workers is that tasks can be completed at faster rates. We observe that this trend holds good across all the task types considered, in case of explicit training of crowd workers. Depending on the type

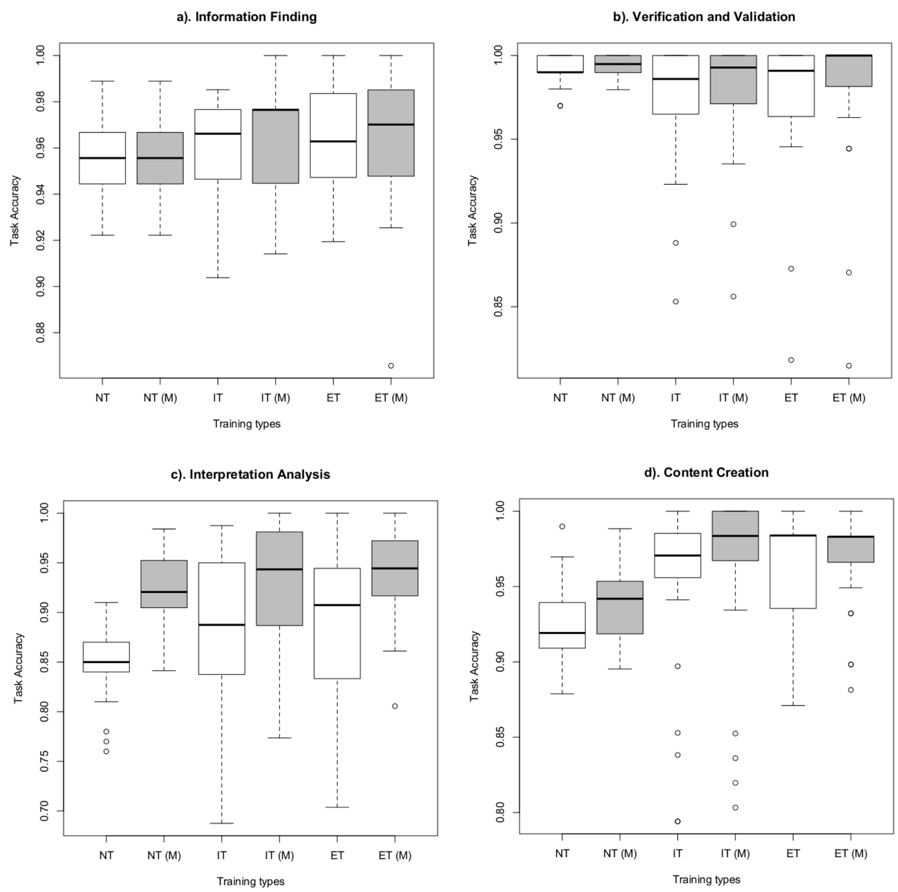


Fig. 7: Overall accuracy of tasks in different training settings in the presence and absence of malicious workers. '(M)' indicates the absence of malicious workers.

of the task, this can help task administrators reduce costs with respect to time to a great extent. For example, when crowd workers learn to complete a relatively complex task in the easiest and most adequate manner, they not only find it easier to complete the task successfully, but they also take lesser time to complete the task at hand.

We highlight that training workers on a particular task goes beyond providing comprehensive instructions. In the explicit training process, workers are required to complete a certain number of units while implementing suggested mechanisms. In the implicit training process, workers are prompted with the training suggestions each time they commit an error. Through our results, we show that these aspects help a worker learn the ideal way to complete a particular microtask successfully. Note that in our work we do not evaluate the quality of the training itself that is provided. We believe that the training process is dependent on the task, and there may be multiple training options for a given task. In the future, we will investigate the research question regarding different training types, durations and their effect on the performance of crowd workers for specific microstasks.

Due to the existing reputation systems in crowdsourcing platforms (such as the *trust-score* in CrowdFlower, we believe that many crowd workers are highly motivated to perform accurately in order to maintain a good reputation or *trust-score*. Previous works have shown that honest workers are also prone to providing inaccurate responses due to their inattentiveness, boredom [5], or in some cases poor instructions and task design [14]. By training the crowd workers, we can help the honest workers in creating and maintaing a good reputation. We observe that the number of *elite workers*, i.e., workers with a 100% task accuracy, increases across different types of tasks due to implicit and/or explicit training.

Interestingly, we note that the training processes discourage malicious workers. It has been established by prior works that malicious workers are mostly attracted to tasks which can be easily completed in short periods of time [3]. From our findings presented in the previous section, we infer that the training processes, both implicit and explicit disrupt a malicious worker from achieving his goal to game the microtask quickly. Such reduction in the fraction of malicious workers, results in a gain in the costs-benifit ratio with respect to the workers responses. In this way, requesters can increase the quantity of useful responses obtained from workers for the costs incurred.

Finally, we address the issue of overheads due to training. Intuitively, the overheads arise primarily due to the training process in terms of time and costs (i.e., monetary compensation). However, we have already established the fact that training crowd workers can actually shorten the time required for task completion. We believe that depending on the crowdsourcing platform being used, the monetary compensation for the additional training processes is either negligible, or worth the minor additional investment.

Although we present the geographical distribution of the workers in our microtasks, we refrain from drawing conclusions with respect to the correlation between specific countries and the types of task. This is due to the fact that the geographical distributions may be skewed by factors such as the time of the day at which the microtasks were deployed on the platform, and the absence of a controlled environment for worker participation. In this regard, more experiments under specific settings are necessary, which is beyond the scope of this work.

6 Conclusions and Future Work

In this paper, we have carried out a comprehensive investigation on the effects of training processes on the performance of workers, in different types of microtasks. We have considered all the applicable types of top-level microtask types according to the taxonomy proposed in our previous work [4].

We delve into two different types of crowd training setups; *implicit* and *explicit* training. Here, we treat crowd workers as *learners* in an atypical learning environment that is characterized by a short learning phase and immediately preceeds the application of learned concepts. Workers learn independently and exploit the training in order to complete the microtasks more efficiently.

Through our experimental results, we have identified that training workers on the specific tasks at hand during their participation in the microtasks, can lead to substantial improvements in their performance (**#RQ1**). In addition, we study the effect of particular kinds of training (implicit/explicit) to assess the optimal kind of training process, in comparison to the cases of 'no-training'. We found that explicit training yields the biggest improvements in the individual performance of workers as well as the overall quality of the results produced (**#RQ2**). Next, we studied how implicit and explicit training effect tasks of different types. We note a trend where explicit training results in the highest improvements across different types of microtasks, followed by implicit training (**#RQ3**). Finally, we studied the effect of training on several aspects such as elite workers (who complete tasks with a 100% accuracy), malicious workers (who attempt to game tasks to avail the monetary compensation easily and quickly), the task completion time, and the overall accuracy of the tasks. We find evidence of a positive impact in all these cases. Training enables workers to complete tasks in shorter time spans, deters malicious workers, and improves the overall task accuracy.

As a part of imminent future work, we plan to model, quantify, and investigate the ideal 'amount of training' that results in high performance gains. We also plan to experiment with varying task lengths, to identify the optimal lengths that render tasks suitable for crowd training. In addition, we intend to investigate other modes of training processes such as a combination of implicit and explicit training.

Acknowledgements. This work has been carried out partially in the context of the DURAARK project, funded by the European Commission within the 7th Framework Programme (Grant Agreement no: 600908).

References

1. S. Do-Lenh, P. Jermann, S. Cuendet, G. Zufferey, and P. Dillenbourg. Task performance vs. learning outcomes: A study of a tangible user interface in the classroom. In *Sustaining TEL: From Innovation to Learning and Practice - 5th European Conference on Technology Enhanced Learning, EC-TEL 2010, Barcelona, Spain, September 28 - October 1, 2010. Proceedings*, pages 78–92, 2010.
2. S. Dow, E. Gerber, and A. Wong. A pilot study of using crowds in the classroom. In *2013 ACM SIGCHI Conference on Human Factors in Computing Systems, CHI '13, Paris, France, April 27 - May 2, 2013*, pages 227–236, 2013.
3. C. Eickhoff and A. P. de Vries. Increasing cheat robustness of crowdsourcing tasks. *Information retrieval*, 16(2):121–137, 2013.
4. U. Gadiraju, R. Kawase, and S. Dietze. A taxonomy of microtasks on the web. In *25th ACM Conference on Hypertext and Social Media, HT '14, Santiago, Chile, September 1-4, 2014*, pages 218–223, 2014.
5. U. Gadiraju, R. Kawase, S. Dietze, and G. Demartini. Understanding malicious behavior in crowdsourcing platforms: The case of online surveys. In *Proceedings of CHI'15, CHI Conference on Human Factors in Computing Systems*, 2015.
6. S. A. Goldman and M. J. Kearns. On the complexity of teaching. In *Proceedings of the Fourth Annual Workshop on Computational Learning Theory*, COLT '91, pages 303–314, San Francisco, CA, USA, 1991. Morgan Kaufmann Publishers Inc.

7. P. G. Ipeirotis, F. Provost, and J. Wang. Quality management on amazon mechanical turk. In *Proceedings of the ACM SIGKDD workshop on human computation*, pages 64–67. ACM, 2010.
8. D. R. Karger, S. Oh, and D. Shah. Iterative learning for reliable crowdsourcing systems. In *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011. Proceedings of a meeting held 12-14 December 2011, Granada, Spain.*, pages 1953–1961, 2011.
9. A. Kobren, C. H. Tan, P. G. Ipeirotis, and E. Gabrilovich. Getting more for less: Optimized crowdsourcing with dynamic tasks and goals. In *Proceedings of the 24th International Conference on World Wide Web, WWW 2015, Florence, Italy, May 18-22, 2015*, pages 592–602, 2015.
10. J. Le, A. Edmonds, V. Hester, and L. Biewald. Ensuring quality in crowdsourced search relevance evaluation: The effects of training question distribution. In *SIGIR 2010 workshop on crowdsourcing for search evaluation*, pages 21–26, 2010.
11. C. H. Lin, Mausam, and D. S. Weld. Crowdsourcing control: Moving beyond multiple choice. In *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence, Catalina Island, CA, USA, August 14-18, 2012*, pages 491–500, 2012.
12. C. C. Marshall and F. M. Shipman. Experiences surveying the crowd: Reflections on methods, participation, and reliability. In *Proceedings of the 5th Annual ACM Web Science Conference*, pages 234–243. ACM, 2013.
13. D. Oleson, A. Sorokin, G. P. Laughlin, V. Hester, J. Le, and L. Biewald. Programmatic gold: Targeted and scalable quality assurance in crowdsourcing. *Human computation*, 11(11), 2011.
14. J. Rogstadius, V. Kostakos, A. Kittur, B. Smus, J. Laredo, and M. Vukovic. An assessment of intrinsic and extrinsic motivation on task performance in crowdsourcing markets. In *ICWSM*, 2011.
15. A. Singla, I. Bogunovic, G. Bartók, A. Karbasi, and A. Krause. On actively teaching the crowd to classify. In *NIPS Workshop on Data Driven Education*, 2013.
16. J. Vuurens, A. P. de Vries, and C. Eickhoff. How much spam can you take? an analysis of crowdsourcing results to increase accuracy. In *Proc. ACM SIGIR Workshop on Crowdsourcing for Information Retrieval (CIR11)*, pages 21–26, 2011.
17. P. Welinder, S. Branson, S. Belongie, and P. Perona. The multidimensional wisdom of crowds. In *Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems 2010. Proceedings of a meeting held 6-9 December 2010, Vancouver, British Columbia, Canada.*, pages 2424–2432, 2010.
18. J. Whitehill, P. Ruvolo, T. Wu, J. Bergsma, and J. R. Movellan. Whose vote should count more: Optimal integration of labels from labelers of unknown expertise. In *Advances in Neural Information Processing Systems 22: 23rd Annual Conference on Neural Information Processing Systems 2009. Proceedings of a meeting held 7-10 December 2009, Vancouver, British Columbia, Canada.*, pages 2035–2043, 2009.
19. A. Xu, H. Rao, S. P. Dow, and B. P. Bailey. A classroom study of using crowd feedback in the iterative design process. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW 2015, Vancouver, BC, Canada, March 14 - 18, 2015*, pages 1637–1648, 2015.
20. M. Zook, M. Graham, T. Shelton, and S. Gorman. Volunteered geographic information and crowdsourcing disaster relief: a case study of the haitian earthquake. *World Medical & Health Policy*, 2(2):7–33, 2010.